

140B Statistical Physics

Part 1. Statistical Ensembles

Information Theory

■ Probability Theory

■ Why Information Theory?

We often think of physics as the quest of the ultimate truth of our universe, but is it really the case?

- Both humans and artificial intelligence (AI) face a fundamental limitation: our **computational power** (the ability to process **information**) is *finite*.

How do we use this limited computational power to comprehend an infinite universe?

This is where **Statistical Mechanics** (StatMech) emerges,

- *not* with a thirst for *exhaustive knowledge*,
- but with a strategy of *selective ignorance*.

Unlike other branches of physics that strive to know more, StatMech plays a different game: it focuses on *what we can afford to forget!*

- It teaches us to *disregard* as many *details* as possible, concentrating our limited computational power on the most crucial **information**.
- This seemingly *passive* approach is our most *powerful* tool to understand **many-body systems**, allowing us to unearth the *patterns* and *regularities* hidden behind the world of *uncertainty* and *randomness*.
- In some sense, **thermodynamics** is the dynamics of *ignorance* (i.e. lack of information, i.e. entropy).

This is why our journey of StatMech will start with **information** theory.

■ Random Variable

In statistical physics, a system can take different states at random:

- The state of a spin (a classical bit): $s = \uparrow, \downarrow$.
- The state of a classical particle: (\mathbf{x}, \mathbf{p}) , $\mathbf{x} \in \mathbb{R}^3$ - position, $\mathbf{p} \in \mathbb{R}^3$ - momentum.
- The state of a quantum harmonic oscillator: $n \in \mathbb{N}$ - energy level.

- The state of an image: $x \in \mathbb{R}^{W \times H \times 3}$ - RGB values of each pixel in a $W \times H$ image.

Message: the **state variable** does not need to be a number, it can be as a vector, a tensor, or a field.

The mathematical language to describe a system is a **random variable** X , which is a variable that can take *random* values x in a set $\mathcal{X} = \{x_1, x_2, \dots\}$ of all *possible values*.

Question: if X = a random image of dog, what are its possible states (values)?



■ Probability

Each **random variable** X also comes with an *assignment* of a **probability** $p(x)$ to each value $x \in \mathcal{X}$, quantifying the degree of *belief* or *certainty* to observe X taking on the value x .

- Probability assignment describes our current state of **knowledge** about the random variable. (It can be subjective.)
- The probability assignment will be *updated*, if our knowledge is *changed* by **observations** (providing new evidence).

Example:

Tossing a coin. x = head, tail.

- **Prior probability** *before* observation

$$\begin{array}{rcc}
 x & \text{head} & \text{tail} \\
 p(x) & \frac{1}{2} & \frac{1}{2}
 \end{array} \tag{1}$$

- **Posterior probability** *after* observing head up.

$$\begin{array}{rcc}
 x & \text{head} & \text{tail} \\
 p(x) & 1 & 0
 \end{array} \tag{2}$$

Observation removes uncertainty and provides **information**.

Properties:

- Positivity:

$$p(x) \geq 0.$$

(3)

- Normalization:

$$\sum_{x \in \mathcal{X}} p(x) = 1. \quad (4)$$

■ Expectation Value

The **expectation value** (mean, average, first-moment) of a random variable X :

$$\langle x \rangle = \mathbb{E}[X] = \sum_{x \in \mathcal{X}} x p(x). \quad (5)$$

Expectation value also can be defined for a *function* of the random variable $f(X)$,

$$\langle f(x) \rangle = \mathbb{E}[f(X)] = \sum_{x \in \mathcal{X}} f(x) p(x). \quad (6)$$

Properties:

- Double expectation

$$\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]. \quad (7)$$

- Linearity (for two random variables X and Y and a constant α)

$$\begin{aligned} \mathbb{E}[X + Y] &= \mathbb{E}[X] + \mathbb{E}[Y], \\ \mathbb{E}[\alpha X] &= \alpha \mathbb{E}[X]. \end{aligned} \quad (8)$$

More generally, for multiple random variables X_i ($i = 1, 2, \dots$) linearly combined together,

$$\mathbb{E}\left[\sum_i \alpha_i X_i\right] = \sum_i \alpha_i \mathbb{E}[X_i]. \quad (9)$$

■ Variance

The **variance** of a random variable X is the expectation of the squared deviation of X from its mean $\mathbb{E}[X]$.

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \langle x^2 \rangle - \langle x \rangle^2. \quad (10)$$

It is a measure of *fluctuation*, i.e. how far the random variable is spread out from its average.

- $\langle x^2 \rangle$ is the second-moment of X

$$\langle x^2 \rangle = \sum_{x \in \mathcal{X}} x^2 p(x). \quad (11)$$

- Positivity: $\text{var } x \geq 0$ (because $(X - \mathbb{E}[X])^2 \geq 0$).

The square root of variance defines the **standard deviation**

$$\text{std}(X) = \sqrt{\text{var}(X)} = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}. \quad (12)$$

Standard deviation measures the amount of *fluctuation* in *linear* scale of the random variable (i.e. $\text{std}(\alpha X) = \alpha \text{std}(X)$ for any constant α).

■ Information and Entropy

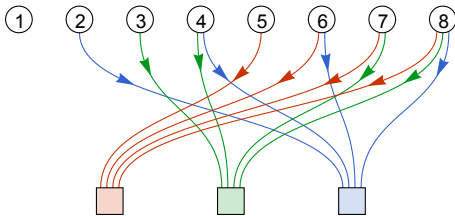
■ Information

Answering every independent **yes-or-no question** provides 1 **bit** of information.

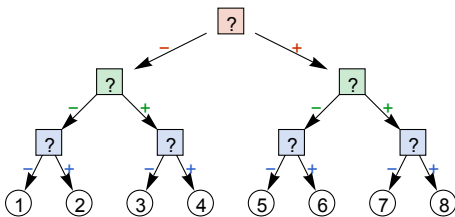
Example: Binary pooled testing

Among eight individuals, there is *one* (and *only one*) contracted COVID-19. How many tests are needed to identify the COVID-19 positive individual?

- Sample collection (encoding scheme)



- Test result analysis (decoding scheme)



In a binary search, answering n independent yes-or-no questions will:

- Identify a unique outcome out of 2^n equally-likely possibilities,
- Collapse the probability from $p = \left(\frac{1}{2}\right)^n$ to $p = 1$ for the observed outcome,
- Provide n bits of information:

$$p = \left(\frac{1}{2}\right)^n \Rightarrow n = -\log_2 p = -\frac{\log p}{\log 2}. \tag{13}$$

$\log 2 = 1$ bit is treated as an **information unit**.

The amount of **information** I gained from the *observation* of a **probability** p outcome is

$$I = -\log p. \text{ (for a particular outcome)}$$

(14)

Example: average information

Observing a random variable X with the following (prior) probability

$$\begin{array}{cccc}
 x & a & b & c & d \\
 p(x) & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8}
 \end{array} \cdot \tag{15}$$

- Information gain can be different for different observation outcomes

$$\begin{aligned}
 I(a) &= -\log(1/2) = \log 2 = 1 \text{ bit}, \\
 I(b) &= -\log(1/4) = 2 \log 2 = 2 \text{ bit}, \\
 I(c) &= -\log(1/8) = 3 \log 2 = 3 \text{ bit}, \\
 I(d) &= -\log(1/8) = 3 \log 2 = 3 \text{ bit}.
 \end{aligned} \tag{16}$$

- However, different outcome happens with different probability. What is *average* amount of information that we can obtain from observing X (regardless of its outcome)?

$$\begin{aligned}
 I(X) &= I(a) p(a) + I(b) p(b) + I(c) p(c) + I(d) p(d) \\
 &= \left(1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 3 \times \frac{1}{8} \right) \text{bit} \\
 &= 1.75 \text{ bit}.
 \end{aligned} \tag{17}$$

In conclusion, given a random variable X , the *expected information* gained from a *full observation* of X is

$$I(X) = -\langle \log p(x) \rangle = -\sum_{x \in \mathcal{X}} p(x) \log p(x). \tag{18}$$

■ Entropy

The **entropy** measures the *uncertainty* (lack of information, ignorance) remained in a random variable X , determined by the *probability distribution* $p(x)$,

$$S(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x). \tag{19}$$

- Entropy is always non-negative (follows from $0 \leq p(x) \leq 1$)

$$S(X) \geq 0. \tag{20}$$

- $S(X) = 0$ means the value of X is known for certain (no randomness).
- Large $S(X)$ indicates large uncertainty in X .
- Entropy can be changed by *observation*, as observation can remove/reduce uncertainty from a random variable.

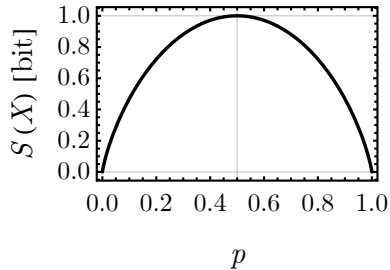
Example:

- A binary random variable X (with $\mathcal{X} = \{\text{false}, \text{true}\}$)

$$\begin{aligned} p(\text{false}) &= 1 - p, \\ p(\text{true}) &= p, \end{aligned} \tag{21}$$

where $0 \leq p \leq 1$. Entropy of X

$$S(X) = -p \log p - (1 - p) \log (1 - p). \tag{22}$$



- $S(X) = 0$ when $p = 0$ ($X = \text{false}$ for sure) or $p = 1$ ($X = \text{true}$ for sure).
- $S(X)$ is maximized at $p = 1/2$, where X is most uncertain. The maximum entropy of a binary random variable is 1 bit.

□ The Unit of Entropy

In statistical physics, the thermal entropy (heat-temperature ratio) is defined with an additional factor k_B ,

$$S(X) = -k_B \sum_{x \in \mathcal{X}} p(x) \log p(x). \tag{23}$$

$k_B = 1.38065 \times 10^{-23} \text{ J/K}$ is the **Boltzmann constant**, a conversion constant between the *temperature* unit and the *energy* unit.

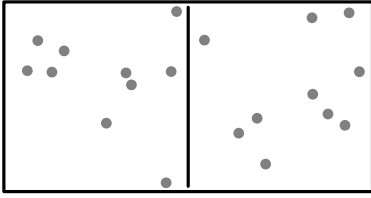
- In **natural units** (Planck units), $k_B = \hbar = c = G = 1$.

Field of physics	const.	appeared in
Statistical mechanics	k_B	$E = k_B T$
Quantum mechanics	\hbar	$E = \hbar \omega$
Special relativity	c	$E = c p$
General relativity	G	$E = -G \frac{M m}{r}$

Fundamental constants are introduced to covert everything to energy.

□ The Epistemological Nature of Entropy

Imagine a box partitioned into two sections. Each half have roughly the same amount of gray balls all bouncing around. [1]



[1] Ruben Verresen. Does entropy depend on observer? Physics StackExchange.

- Remove, wait, and re-insert the partition - not much changed - there has been no increase of entropy:

$$\Delta S = 0. \tag{24}$$

- However, it turns out we were color blind. The balls are red and green, and initially all red (green) balls are on the left (right). Mixing will increase the entropy by (1 bit for each ball)

$$\Delta S = N \log 2. \tag{25}$$

Which statement is correct?

- Both are correct.
- Entropy is *not* a property of the physical system. It depends on what we can (or choose to) measure, and how much details we are able (or willing) to look into.
- **Information** can be used to extract **work**.
 - Replace the partition with a semi-permeable membrane that only allows red balls to pass through (i.e., the membrane is not color-blind). It will experience an imbalanced pressure from the green side, known as the **osmotic pressure**, which can be used to do work.
 - To us, as color-blind observers, this seems to *violate* the **second law of thermodynamics**, that it is impossible to extract work from a maximal entropy system.
 - However, observing such a work done would force us to reconsider our limitations, and to acknowledge that there must be some underlying difference among the balls, even if we cannot directly perceive it!

■ Statistical Inference

■ Maximum Entropy Estimation

Maximum entropy (MaxEnt) estimation is a cornerstone in the realm of *statistical inference*, offering a principle for estimating *probability distributions* of random variables. [2]

The probability distribution of a random variable X should be assigned to *maximize* the *entropy* of X , within the bounds of our existing knowledge about X .

- **Entropy** serves as a metric for our *ignorance*. By maximizing entropy, we are maximizing our ignorance in a controlled way.
- We should be *honest about our ignorance*. MaxEnt insists that our probabilistic beliefs should be calibrated such that they do not pretend a level of knowledge that exceeds what our evidence or hypotheses justify.

[2] E. T. Jaynes. Why do we Stand on Maximum Entropy. <https://bayes.wustl.edu/etj/articles/stand.on.entropy.pdf>

Example: uniform distribution

Consider a random variable X with Ω possible values. Without further assumptions, the maximum entropy probability distribution is the **uniform (equal probability) distribution**:

$$\forall x \in \mathcal{X} : p(x) = \frac{1}{\Omega}, \quad (26)$$

which has entropy

$$S(X) = \log \Omega. \quad (27)$$

Exc
1

Prove that the equal probability distribution is maximal entropy.

■ Gibbs Distribution

Let X be a random variable, with **constraints** on the *expectation values* for a series of functions of x : $f_1(x)$, $f_2(x)$, ...

$$\langle f_a(x) \rangle = \sum_{x \in \mathcal{X}} f_a(x) p(x), \quad (28)$$

the MaxEnt estimation can be formulated as a **constrained optimization problem**:

$$\begin{aligned} \max_{p(x)} S(X) &= - \sum_{x \in \mathcal{X}} p(x) \log p(x), \\ \text{s.t. } \forall a : \sum_{x \in \mathcal{X}} f_a(x) p(x) &= \langle f_a(x) \rangle. \end{aligned} \quad (29)$$

the optimal probability distribution is the **Gibbs distribution** [3]

$$p(x) = \frac{1}{Z[\lambda]} \exp\left(- \sum_a \lambda_a f_a(x)\right), \quad (30)$$

Exc
2

Use the Lagrangian multiplier method to show that Eq. (30) is the optimal solution of Eq. (29).

[3] E. T. Jaynes. Information Theory and Statistical Mechanics. <https://bayes.wustl.edu/etj/articles/theory.1.pdf>

- $p(x)$ takes an **exponential form**, where the exponent is a linear combination of all *constraint functions* $f_a(x)$ by corresponding *Lagrangian multipliers* λ_a

$$p(x) \propto \exp\left(-\sum_a \lambda_a f_a(x)\right). \quad (31)$$

- $Z[\lambda]$ is called the **partition function**,

$$Z[\lambda] = \sum_{x \in \mathcal{X}} \exp\left(-\sum_a \lambda_a f_a(x)\right), \quad (32)$$

serving as the normalization factor to ensure $\sum_{x \in \mathcal{X}} p(x) = 1$.

- λ_a are *Lagrangian multipliers*, which should be determined by solving the corresponding **constraint equation**

$$-\frac{\partial}{\partial \lambda_a} \log Z[\lambda] = \langle f_a(x) \rangle. \quad (33)$$

Exc
3

Derive Eq. (33).

- The **entropy** of the optimal distribution Eq. (30) is

$$\begin{aligned} S(X) &= \log Z[\lambda] + \sum_a \lambda_a \langle f_a(x) \rangle \\ &= \left(1 - \sum_a \frac{\partial}{\partial \log \lambda_a}\right) \log Z[\lambda], \end{aligned} \quad (34)$$

which has been maximal under the constraints of $\langle f_a(x) \rangle$.

Exc
4

Derive Eq. (34).

Consider a six-sided dice (with face numbered 1 to 6). However, the dice may not be fair. You are given the knowledge that the expectation value of a roll is 4.5. Your task is to determine the probability distribution of outcome numbers using the maximal entropy estimation.

(i) Define the random variable in the problem. Given its expectation value, formulate the constraint for the probabilities.

(ii) Use the maximal entropy principle to set up the constrained optimization problem. Solve the problem by the Lagrangian multiplier method or by a computer program.

(iii) Make a table of the probability distributions you found (numerical values), like the following:

x	1	2	3	4	5	6
$p(x)$	□	□	□	□	□	□

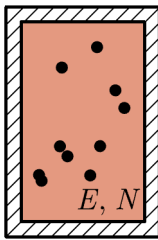
HW
1

Statistical Ensembles

■ Microcanonical Ensemble

■ Probability Distribution

The **microcanonical ensemble** describes **isolated (closed) systems**, with fixed *energy* E and *number of particles* N .



- All accessible **microstates** $x \in \mathcal{X}_{E,N}$ of the system have the same energy E and particle number N . The ensemble of all such systems belongs to the same **macrostate** labeled by (E, N) .
- Without further constraints, the MaxEnt estimation asserts that the probability distribution is **uniform**, i.e., every microstate is equally likely

$$p(x) = \frac{1}{\Omega(E, N)}, \quad (35)$$

where

- $\Omega(E, N)$ is the **number of microstates** in $\mathcal{X}_{E,N}$ (the set of microstates with energy E and particle number N),

$$\Omega(E, N) = \sum_{x \in \mathcal{X}_{E,N}} 1. \quad (36)$$

In microcanonical ensemble, $\Omega(E, N)$ serves the role of the **partition function**.

■ Thermodynamics

The **entropy** is given by [recall the entropy formula Eq. (34)]

$$S(E, N) = \log \Omega(E, N). \quad (37)$$

- The entropy function enables us to define **temperature** T and **chemical potential** μ through its derivatives, [accept these definitions as historical conventions for now]

$$\left(\frac{\partial S}{\partial E}\right)_N = \frac{1}{T},$$

$$\left(\frac{\partial S}{\partial N}\right)_E = -\frac{\mu}{T}.$$
(38)

- In differential form,

$$dS = \frac{1}{T} dE - \frac{\mu}{T} dN,$$
(39)

which can be written in a more familiar form as the **thermodynamic identity** for energy

$$dE = T dS + \mu dN.$$

(40)

- This is also known as the **first law of thermodynamics**, which states that the energy change dE is a sum of heat $T dS$ and work μdN .
- The conjugate pair (μ, N) can be replaced by other pairs in other thermodynamic problems, but the underlying principles are the same. [See the section Conjugate Pairs of Variables for more examples.]

□ Equilibrium Condition

Why are *temperature* and *chemical potential* defined in this way (as in Eq. (38))? This has to do with what we mean by *thermal* and *chemical equilibrium*.

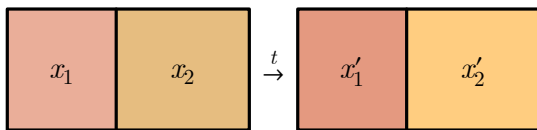
Equilibrium: probability distribution does not evolve in time.

Consider two initially isolated systems.

- The probability to observe them in microstates (x_1, x_2) is

$$p_0(x_1, x_2) = p(x_1) p(x_2).$$
(41)

- Imagining bringing the two systems in contact for some time t . They jointly evolve into new microstates:



Assuming the **time evolution** is *deterministic*, the probability to observe (x'_1, x'_2) after the contact is

$$p_t(x'_1, x'_2) = p_0(x_1, x_2) = p(x_1) p(x_2).$$
(42)

- If the two systems were in equilibrium, their joint probability distribution should be time independent, meaning

$$p_t(x'_1, x'_2) = p_0(x'_1, x'_2) = p(x'_1) p(x'_2). \quad (43)$$

So the **equilibrium condition** (at the probability level) is

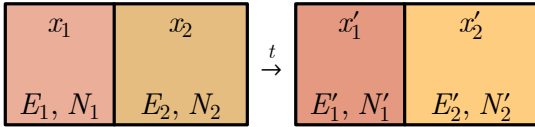
$$p(x_1) p(x_2) = p(x'_1) p(x'_2), \quad (44)$$

for any (x'_1, x'_2) related to (x_1, x_2) by time evolution.

What is its implication for microcanonical ensembles?

For microcanonical ensembles, $p(x)$ only depends on the energy E and particle number N of the microstate x

$$p(x) = \frac{1}{\Omega(E, N)}. \quad (45)$$



The equilibrium condition Eq. (44) implies

$$\frac{1}{\Omega(E_1, N_1)} \frac{1}{\Omega(E_2, N_2)} = \frac{1}{\Omega(E'_1, N'_1)} \frac{1}{\Omega(E'_2, N'_2)}, \quad (46)$$

Taking logarithm on both side, noting $S = \log \Omega$, Eq. (46) becomes

$$S(E_1, N_1) + S(E_2, N_2) = S(E'_1, N'_1) + S(E'_2, N'_2), \quad (47)$$

meaning that the **total entropy** will not evolve in time, if two systems have reached equilibrium. Note that the *energy* and *particle number* are also time invariant by conservation laws,

$$\begin{aligned} E_1 + E_2 &= E'_1 + E'_2, \\ N_1 + N_2 &= N'_1 + N'_2. \end{aligned} \quad (48)$$

These can be summarized in differential forms as (under infinitesimal evolution of time dt)

$$\begin{aligned} dE_1 + dE_2 &= 0, \\ dN_1 + dN_2 &= 0, \\ dS(E_1, N_1) + dS(E_2, N_2) & \\ &= \left(\frac{\partial S}{\partial E_1} \right)_{N_1} dE_1 + \left(\frac{\partial S}{\partial N_1} \right)_{E_1} dN_1 + \left(\frac{\partial S}{\partial E_2} \right)_{N_2} dE_2 + \left(\frac{\partial S}{\partial N_2} \right)_{E_2} dN_2 \\ &= 0. \end{aligned} \quad (49)$$

Result in two **equilibrium conditions** (at the thermodynamic level)

Exc
5

Derive the following two equilibrium conditions from Eq. (49).

- The **thermal equilibrium** condition

$$\frac{1}{T_1} = \left(\frac{\partial S}{\partial E_1} \right)_{N_1} = \left(\frac{\partial S}{\partial E_2} \right)_{N_2} = \frac{1}{T_2}, \quad (50)$$

or equivalently

$$T_1 = T_2, \quad (51)$$

which defines what should be called **temperature**.

- The **chemical equilibrium** condition

$$-\frac{\mu_1}{T_1} = \left(\frac{\partial S}{\partial N_1} \right)_{E_1} = \left(\frac{\partial S}{\partial N_2} \right)_{E_2} = -\frac{\mu_2}{T_2}, \quad (52)$$

or equivalently (assuming $T_1 = T_2$ is already achieved)

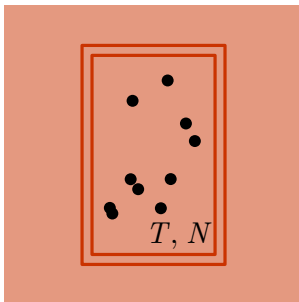
$$\mu_1 = \mu_2, \quad (53)$$

which defines what should be called **chemical potential**.

■ Canonical Ensemble

■ Probability Distribution

The **canonical ensemble** describes **isothermal systems** (thermally open and particle-wise closed), in thermal equilibrium with a heat bath, with fixed *temperature* T and *number of particles* N .



- *Microstates* $x \in \mathcal{X}_N$ of the system still have the same particle number N , but can take all different energies.
- An **energy function** $E(x)$ is therefore introduced to evaluate the energy associated with each state x .
- One can still talk about the **energy expectation value** for the ensemble

$$E = \langle E(x) \rangle = \sum_{x \in \mathcal{X}_N} E(x) p(x). \quad (54)$$

Given temperature T , the ensemble is expected to have an energy expectation value $\langle E(x) \rangle$

that matches the *macroscopic* energy E .

- Given the **energy constraint** $\langle E(x) \rangle = E$, the MaxEnt estimation asserts that the probability distribution should be [recall Eq. (30)]

$$p(x) = \frac{1}{Z(\beta, N)} e^{-\beta E(x)}, \quad (55)$$

where

- The partition function $Z(\beta, N)$ is given by

$$Z(\beta, N) = \sum_{x \in \mathcal{X}_N} e^{-\beta E(x)}. \quad (56)$$

- Originally introduced as a *Lagrangian multiplier*, β is found to be related to the **temperature** T by (assuming $k_B = 1$)

$$\beta = \frac{1}{T}, \quad (57)$$

so β is also called the **inverse temperature**.

□ Lagrangian Multipliers

The Lagrangian multiplier β was introduced to enforce the *energy constraint* $\langle E(x) \rangle = E$, so it should be determined by solving the constraint equation, according to Eq. (33),

$$-\partial_{\lambda_a} \log Z[\lambda] = \langle f_a \rangle,$$

$$E = - \left(\frac{\partial \log Z}{\partial \beta} \right)_N. \quad (58)$$

On the other hand, Eq. (34), $S = \log Z + \sum_a \lambda_a \langle f_a \rangle$, tells us how to compute *entropy*,

$$S = \log Z + \beta E. \quad (59)$$

Take the total derivative on both sides,

$$dS = \beta dE + \left(\frac{\partial \log Z}{\partial N} \right)_\beta dN. \quad (60)$$

**Exc
6**

Derive Eq. (60) using Eq. (58).

Compared with Eq. (39), which defines temperature and chemical potential,

$$dS = \frac{1}{T} dE - \frac{\mu}{T} dN, \quad (61)$$

one concludes that

$$\beta = \frac{1}{T},$$

$$\mu = -T \left(\frac{\partial \log Z}{\partial N} \right)_\beta = -T \left(\frac{\partial \log Z}{\partial N} \right)_T. \quad (62)$$

■ Thermodynamics

The logarithm of partition function $\log Z$ defines the (Helmholtz) **free energy**,

$$F(T, N) = -T \log Z(T, N). \quad (63)$$

- The free energy enables us to compute:
 - Entropy

$$S = - \left(\frac{\partial F}{\partial T} \right)_N, \quad (64)$$

Exc 7 | Derive Eq. (64).

- Chemical potential

$$\mu = \left(\frac{\partial F}{\partial N} \right)_T. \quad (65)$$

Exc 8 | Derive Eq. (65).

- In differential form,

$$dF = -S dT + \mu dN. \quad (66)$$

This is the **thermodynamic identity** for free energy.

Moreover, the free energy is related to the energy by

$$F = E - T S. \quad (67)$$

Exc 9 | Derive Eq. (67) from Eq. (59).

Conversely, the energy can be reconstructed from the free energy

$$E = F - T \left(\frac{\partial F}{\partial T} \right)_N. \quad (68)$$

Statistical physics of a qubit. A qubit is a system with two states (labeled by $x = 0, 1$): a ground state of energy $E(0) = 0$, and an excited state of energy $E(1) = \epsilon$ (where $\epsilon > 0$). Suppose the qubit is in thermal equilibrium with a heat bath at temperature T . (The number of particles can be treated as $N = 1$, and is not relevant to the problem.)

HW
2

(i) Calculate the partition function Z in terms of ϵ and $T = 1/\beta$.

(ii) Derive the expression for the probabilities $p(x)$ of finding the qubit in the ground state ($x = 0$) and the excited state ($x = 1$), respectively.

Let us compute the entropy S in two different ways:

(iii) From the partition function Z , calculate the free energy $F = -T \log Z$, and then obtain $S = -\partial F / \partial T$.

(iv) From the probability distribution $p(x)$, calculate the entropy by definition $S = -\sum_x p(x) \log p(x)$.

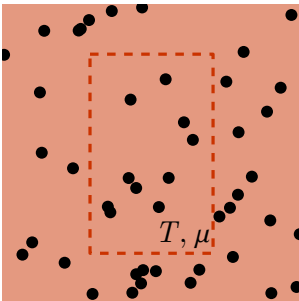
Compare the results obtained in (iii) and (iv) and show that they are the same.

(v) Plot the entropy S as a function of temperature T (taking $\epsilon = 1$ as the energy unit), discuss the behavior at low and high temperatures.

■ Grand Canonical Ensemble

■ Probability Distribution

The **grand canonical ensemble** describes **open systems**, in thermal and chemical equilibrium with a reservoir, at fixed *temperature* T and *chemical potential* μ .



- *Microstates* $x \in \mathcal{X}$ can take all possible states of different energies and particle numbers.
- To evaluate the energy and particle number associated with a particular *microstate* x , the **energy function** $E(x)$ and the **particle number function** $N(x)$ are introduced respectively.
- The energy E and particle number N of the *macrostate* refers to their **expectation values** in grand canonical ensemble

$$\begin{aligned}
E &= \langle E(x) \rangle = \sum_{x \in \mathcal{X}} E(x) p(x), \\
N &= \langle N(x) \rangle = \sum_{x \in \mathcal{X}} N(x) p(x).
\end{aligned}
\tag{69}$$

- Given the **energy constraint** $\langle E(x) \rangle = E$ and the **particle number constraint** $\langle N(x) \rangle = N$, the MaxEnt estimation asserts that the probability distribution should be [recall Eq. (30)]

$$p(x) = \frac{1}{\mathcal{Z}(\beta, \mu)} e^{-\beta(E(x) - \mu N(x))},
\tag{70}$$

where

- The (grand) partition function $\mathcal{Z}(\beta, \mu)$ is given by

$$\mathcal{Z}(\beta, \mu) = \sum_{x \in \mathcal{X}} e^{-\beta(E(x) - \mu N(x))}.
\tag{71}$$

- The *Lagrangian multipliers* are parameterized by the **inverse temperature** $\beta = 1/T$ and the **chemical potential** μ .

□ Lagrangian Multipliers

To figure out the physical meaning of Lagrangian multipliers, first rewrite the partition function Eq. (71) as

$$\mathcal{Z}(\beta, -\gamma/\beta) = \sum_{x \in \mathcal{X}} e^{-\beta E(x) - \gamma N(x)},
\tag{72}$$

such that β and γ can be treated as independent Lagrangian multipliers. According to Eq. (33), $-\partial_{\lambda_a} \log Z[\lambda] = \langle f_a \rangle$, their corresponding constraint equations are

$$\begin{aligned}
E &= - \left(\frac{\partial \log \mathcal{Z}}{\partial \beta} \right)_{\gamma}, \\
N &= - \left(\frac{\partial \log \mathcal{Z}}{\partial \gamma} \right)_{\beta},
\end{aligned}
\tag{73}$$

On the other hand, Eq. (34), $S = \log Z + \sum_a \lambda_a \langle f_a \rangle$, tells us how to compute *entropy*,

$$S = \log \mathcal{Z} + \beta E + \gamma N.
\tag{74}$$

Take the total derivative on both sides,

$$dS = \beta dE + \gamma dN.
\tag{75}$$

Exc 10 | Derive Eq. (75) using Eq. (73).

Compared with Eq. (39), which defines temperature and chemical potential,

$$dS = \frac{1}{T} dE - \frac{\mu}{T} dN, \quad (76)$$

one concludes that

$$\begin{aligned} \beta &= \frac{1}{T}, \\ \gamma &= -\frac{\mu}{T}, \end{aligned} \quad (77)$$

which justify the parametrization of Lagrangian multipliers in Eq. (70) and Eq. (71).

■ Thermodynamics

The logarithm of (grand) partition function $\log \mathcal{Z}$ defines the (grand) **free energy**,

$$\mathcal{F}(T, \mu) = -T \log \mathcal{Z}(1/T, \mu). \quad (78)$$

- Some textbook defines a similar concept, call the “grand potential”: $\Phi(T, \mu) = -\mathcal{F}(T, \mu)$.
- The (grand) free energy enables us to compute:
 - Entropy

$$S = -\left(\frac{\partial \mathcal{F}}{\partial T}\right)_\mu, \quad (79)$$

Exc 11 | Derive Eq. (79).

- Particle number

$$N = -\left(\frac{\partial \mathcal{F}}{\partial \mu}\right)_T, \quad (80)$$

Exc 12 | Derive Eq. (80).

- In differential form,

$$d\mathcal{F} = -S dT - N d\mu. \quad (84)$$

This is the **thermodynamic identity** for (grand) free energy.

Moreover, the (grand) free energy is related to the energy by

$$\mathcal{F} = E - T S - \mu N. \quad (85)$$

Exc 13 | Derive Eq. (85) from Eq. (74).

Conversely, the energy can be reconstructed from the (grand) free energy.

$$E = \mathfrak{F} - T \left(\frac{\partial \mathfrak{F}}{\partial T} \right)_\mu - \mu \left(\frac{\partial \mathfrak{F}}{\partial \mu} \right)_T. \quad (86)$$

■ Summary

■ Statistical Ensembles

Comparison of the three primary statistical ensembles:

Ensemble	Microcanonical	Canonical	Grand Canonical
Microstates	$\mathcal{X}_{E,N}$	$\subseteq \mathcal{X}_N$	$\subseteq \mathcal{X}$
Macrostate	(E, N)	(T, N)	(T, μ)
Probability	$p = \frac{1}{\Omega}$	$p = \frac{1}{Z} e^{-\beta E}$	$p = \frac{1}{\mathcal{Z}} e^{-\beta(E-\mu N)}$
Partition function	$\Omega = \sum 1$	$Z = \sum e^{-\beta E}$	$\mathcal{Z} = \sum e^{-\beta(E-\mu N)}$
Free energy	none	$F = -T \log Z$	$\mathfrak{F} = -T \log \mathcal{Z}$
Entropy	$S = \log \Omega$	$S = -\left(\frac{\partial F}{\partial T} \right)_N$	$S = -\left(\frac{\partial \mathfrak{F}}{\partial T} \right)_\mu$
Energy	E	$E = F + T S$	$E = \mathfrak{F} + T S + \mu N$
Particle number	N	N	$N = -\left(\frac{\partial \mathfrak{F}}{\partial \mu} \right)_T$
Temperature	$T = \left(\frac{\partial E}{\partial S} \right)_N$	T	T
Chemical potential	$\mu = \left(\frac{\partial E}{\partial N} \right)_S$	$\mu = \left(\frac{\partial F}{\partial N} \right)_T$	μ
Thermodynamics	$dE = T dS + \mu dN$	$dF = -S dT + \mu dN$	$d\mathfrak{F} = -S dT - N d\mu$

■ Conjugate Pair of Variables

In thermodynamics, a **conjugate pair** of variables is a pair of

- **Intensive variable** (like T, μ) to be *balanced* at equilibrium,
- **Extensive variable** (like E, N) to be *transferred* in order to reach the equilibrium, such that they multiplies to an *energy* unit.

We have only focused on (T, S) and (μ, N) pairs in the above discussion. However, the principle of statistical mechanics can be applied to much broader systems:

Process / System	Conjugate pair of variables		Contribution to energy (dE)
	Intensive variable	Extensive variable	
Heat transfer	Temperature (T)	Entropy (S)	$+ T dS$
Particle transfer	Chemical potential (μ)	Particle number (N)	$+ \mu dN$
Charge transfer	Electric potential (\mathcal{E})	Charge (Q)	$+ \mathcal{E} dQ$
Gas, Liquid	Pressure (P)	Volume (V)	$- P dV$

Membrane	Surface tension (γ)	Area (A)	$-\gamma dA$
Elastic material	Stree tensor (σ)	Strain tensor (ϵ)	$+\text{Tr } \sigma d\epsilon$
Dielectric material	Electric field (\mathbf{D})	Electric polarization (\mathbf{P})	$-\frac{1}{\epsilon} \mathbf{D} \cdot d\mathbf{P}$
Magnetic material	Magnetic field (\mathbf{B})	Magnetization (\mathbf{M})	$-\mathbf{B} \cdot d\mathbf{M}$
Rotating galaxy	Angular velocity (ω)	Angular momentum (L)	$+\omega dL$

This flexibility allows for a comprehensive modeling of a wide range of physical phenomena.

For example, we want to model a system of

☑ Gas ($-P dV$),

with

☑ Heat transfer ($+T dS$),

☑ Particle transfer ($+\mu dN$).

- The thermodynamics is described by

$$dE = -P dV + T dS + \mu dN, \quad (87)$$

or, in terms of free energies,

$$\begin{aligned} dF &= -P dV - S dT + \mu dN, \\ d\mathcal{F} &= -P dV - S dT - N d\mu. \end{aligned} \quad (88)$$

- This enables us to write done thermodynamic identities, such as

$$\begin{aligned} P &= -\left(\frac{\partial \mathcal{F}}{\partial V}\right)_{T,\mu}, \\ S &= -\left(\frac{\partial \mathcal{F}}{\partial T}\right)_{V,\mu}, \\ N &= -\left(\frac{\partial \mathcal{F}}{\partial \mu}\right)_{V,T}. \end{aligned} \quad (89)$$